Pergamon

Bioorganic & Medicinal Chemistry, Vol. 5, No. 6, pp 1087–1096, 1997

PII: S0968-0896(97)00044-8
0968-0896/97 $17.00 + 0.00

Post-SELEX Combinatorial Optimization of Aptamers

Bruce E. Eaton, Larry Gold,* Brian J. Hicke, Nebojša Janjić, Fiona M. Jucker, David P. Sebesta, Theodore M. Tarasow, Michael C. Willis and Dominic A. Zichi

NeXstar Pharmaceuticals, Inc., 2860 Wilderness Place, Boulder, CO 80301, U.S.A.

Abstract—In vitro selection techniques provide a means of isolating nucleic acid ligands for binding to particular protein targets. Although most aptamers have quite high affinities for their target proteins, it has been shown that post-SELEX modification can result in further enhancement of binding affinity, as well as other desired properties. This has led to the current development of a more systematic approach to aptamer optimization using a combinatorial screening methodology. © 1997 Elsevier Science Ltd.

Introduction

Screening synthetic chemical libraries for a broad spectrum of desired functions has become a powerful tool for drug discovery, catalyst generation, and the study of structure–function relationships. These libraries have been comprised of linear polymers, such as peptides and oligonucleotides, as well as branched small molecules with chemically diverse substituents. The focus of our research has been the isolation of oligonucleotides that are capable of binding to molecular targets with both high affinity and specificity. These ligands, although well suited for their selected function, typically undergo subsequent modification to enhance other desired properties. The development of a combinatorial approach to such modifications is the topic of this work.

SELEX: oligonucleotide combinatorial paradigm

'Systematic evolution of ligands by exponential enrichment' (SELEX) is a protocol in which vast libraries of single-stranded oligonucleotides are screened for desired activities. The first SELEX experiment on single-stranded oligonucleotides was published by Tuerk and Gold.[1] Many combinatorial chemistry paradigms are one-shot screens; that is, the library is not capable of amplification after the first screening. Peptide combinatorial chemistry systems include both nonamplifiable[2,3] and amplifiable[4] versions. In general amplifiable combinatorial chemistry methods allow more thorough screening for extremely rare compounds, uncomplicated by many of the background problems intrinsic to nonamplifiable systems.[5] SELEX is an *oligonucleotide* discovery system; that is, the compounds sought (as drugs, as diagnostic reagents, as catalysts, etc) are oligonucleotides and have been called aptamers.[6] While the first SELEX experiment was performed with an RNA library (using the four natural nucleotide monomers of RNA), it was clear that single-stranded DNA and many modified oligonucleotides also were suitable sources of shape diversity. The idea was that oligonucleotides were not only tapes, but were also shapes, and that sufficiently large oligonucleotide libraries were more representative of 'shape space' than other combinatorial chemistry libraries. SELEX has been used to find a large number of novel catalysts and novel binding reagents; these experiments have been reviewed in detail elsewhere.[7]

The SELEX procedure usually is initiated with a vast library consisting of some 10^{14}–10^{15} randomized oligonucleotide sequences (greater than typical combinatorial chemistry libraries by a factor of 10^{10}!). The molecular diversity is created by randomizing nucleotides, often between two fixed sequences that are used for amplification. For randomization over N nucleotides with four monomers, the sequence space is 4^N. Linear oligonucleotide sequence diversity drives shape diversity, and these libraries yield molecular shapes that were unanticipated by our prior knowledge of nucleic acid structures. No-one knows what fraction of the 10^{15} sequences of single-stranded oligonucleotides, randomized over, for example, 50 nucleotides, have functional shapes. We think that the fraction is substantial and greater than the structured fraction of 10^{15} peptides randomized over a comparable distance, even if those peptides are constrained by, for example, disulfide bonds between cysteines.[8]

The SELEX methodology is outlined in Figure 1. Partitioning is a key step that occurs during the screening. In this step, the oligonucleotide library can be screened for binding to a molecular target or for performing a catalytic function. During a round of SELEX a subset of the initial library that performs the catalytic or binding function is amplified so that the subsequent screening round is performed on only the library members that survived the earlier round. While physical partitioning has been successful, selective amplification can be achieved without physical partitioning by coupling the availability of the fixed sequences for amplification to the event being selected.[9] Partitioning is extremely important, especially during the early rounds, since we have found that many useful

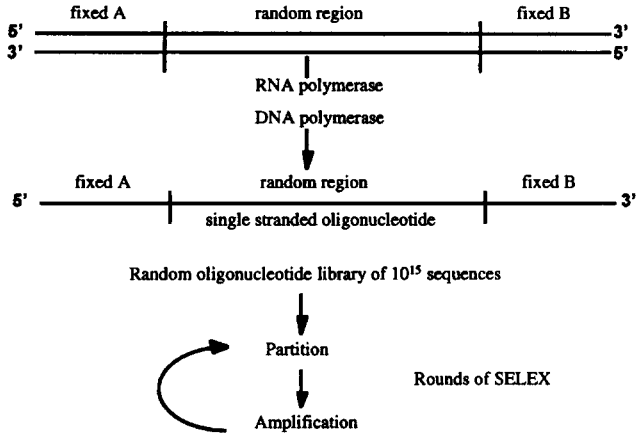

Figure 1. Diagram of the SELEX process. A library comprised of 10^{15} sequences of single-stranded oligonucleotides is prepared. Each sequence contains a random region flanked by fixed sequences, labeled A and B here, that allow for amplification and transcription. A round of SELEX consists of partitioning the library based on function followed by amplification of the selected population. After several rounds the resulting population is cloned and sequenced.

oligonucleotides identified with SELEX are present in the initial library in small fractions, perhaps as low as 10^{-13}.[7]

At this point aptamers for more than 100 protein targets have been isolated, with more than half of the examples published. The weakest affinity reported for any protein target for which the SELEX experiment was done thoroughly is a K_d of about 100 nM.[7] The target of that SELEX, nerve growth factor, is an extremely basic protein and has quite high nonspecific affinity for nucleic acids. As noted by Irvine et al., such proteins would be more difficult targets for SELEX because of a signal-to-noise problem.[5] Oligonucleotides aimed at every other protein target have higher affinities, with K_d values in the range of nanomolar to picomolar. Thus for protein targets SELEX yields compounds with affinities comparable to or better than the affinities of antibodies, especially when one considers that most reported antibody affinities are for bivalent reagents, or interactions between a soluble target and a patch of high density ligand on a matrix, or both! The K_d values reported for SELEX-derived oligonucleotides are true solution affinities, uncomplicated by issues of cooperativity or avidity. Although the focus in this discussion is often on inhibition and antagonists, it can be easily envisaged that, in the case of receptors, antagonists can be converted into agonists.[10,11]

Finally, aptamers for any target protein can be discovered with SELEX regardless of the protein's native function. Regulatory proteins, housekeeping proteins, receptors, growth factors, enzymes, any protein at all can be the subject of a successful SELEX search for a high affinity oligonucleotide ligand.[12] Recent excitement about proteins that bind nucleotides in nature and the capacity of those proteins to perform some regulatory event by binding is only the beginning.[13] SELEX will greatly expand our understanding of protein–nucleic acid interactions.[14,15] All proteins can have high-affinity oligonucleotide ligands, even if those proteins don't interact with nucleotides as part of their native function.

At the completion of the SELEX process the subset of surviving sequences are determined. When more than one sequence is obtained (which is always the goal), consideration of the sequence set shows conserved sequence and structural elements. This kind of comparative sequence analysis has been particularly profound in the study of phylogeny in work pioneered by Woese and his colleagues.[16,17] SELEX of course has no phylogeny[18] but rather is the culling from a vast pool of totally unrelated sequences that were present from the beginning of the selection. Nevertheless, in practice, comparative sequence analysis has been used to identify both sequence and secondary structure conservation to aid in the development of three-dimensional (3-D) models.[19,20] Usually, but not always, the conserved sequences within a ligand family are found within a fixed location in the predicted secondary structures of the ligands. The winning sequences form 3-D structures whose details await actual structural determination.[21–24]

The molecules identified with SELEX can be characterized more fully by post-SELEX experiments in order to define those atomic features of the ligand responsible for the binding activity. Typically only a small subset of nucleotides from a given SELEX-derived sequence is *directly* responsible for high-affinity binding. Truncation experiments have been used to define the ligand boundaries necessary for full binding activity. This results in the identification of nucleotide motifs shorter than the full length ligands, often from 25 to 40 nucleotides in length, and usually missing the fixed sequences utilized during the SELEX rounds for amplification.[25]

Further characterization of the ligand/target complex is possible through chemical probing in the presence and absence of the target molecule, a set of techniques well known as footprinting and binding interference. In these experiments, chemical reactivity depends upon solvent accessibility of specific functional groups on the ligand, targeted by a variety of chemical probes.[26,27] Differential chemical reactivity with respect to bound and unbound molecules identifies those aptamer atoms in direct contact with (or at least in proximity to) the protein target. Alternatively, chemical modification in the absence of the target followed by selection can reveal positions where modification interferes with function. In addition, chemical- and photo-crosslinking experiments can identify protein aptamer contacts. With a known protein 3-D structure, such crosslinking data can be useful for modeling the complex, identifying an array of potential side chain and aptamer contacts to be exploited in library design (see below). Figure 2 displays the results of such crosslinking and chemical probing experiments on an aptamer that binds to human L-selectin.[28] The chemical probing data suggest two faces of the aptamer that interact with L-selectin,

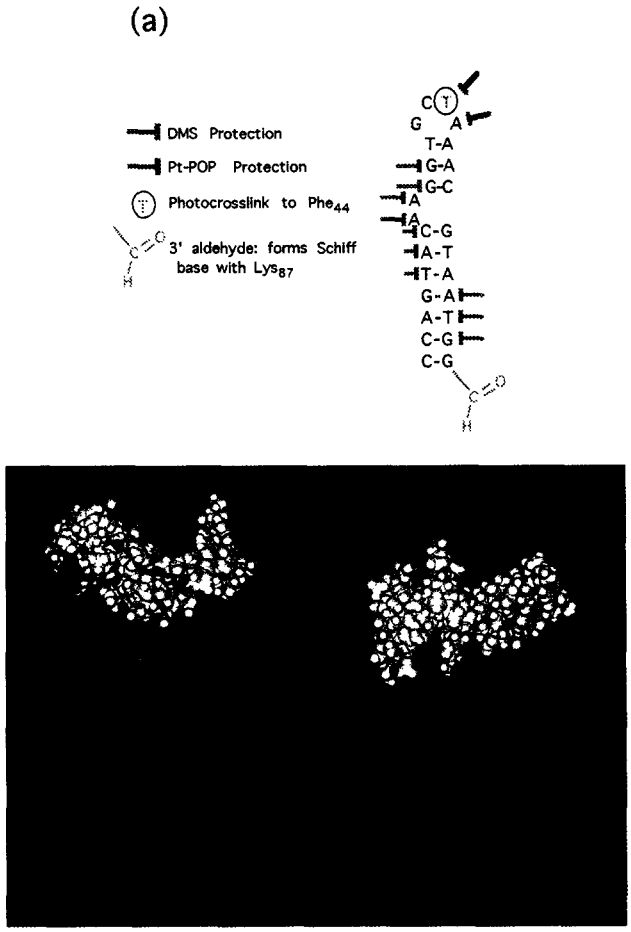

Figure 2. Mapping protein contacts to an aptamer. (a) The secondary structure of an ssDNA aptamer, a 26 nucleotide stem-loop with an internal asymmetric AA bulge, that binds to L-selectin is displayed along with chemical probing and crosslinking results. The chemical protection and crosslinking data are schematically presented on the proposed secondary structure of the DNA aptamer. (b) A 3-D model of the aptamer (white) is shown, with the footprinting sites and crosslinking groups colored as in (a), above the protein target L-selectin (blue) with crosslinking residues colored red. (c) A model for the aptamer/protein complex was generated consistent with the photocrosslinking data, Phe44 stacked on T15, and reductive amination data, Lys87 near a 3′-aldehyde moiety. Aptamer binding is also EDTA sensitive, suggesting a possible role for the Ca^{2+} (orange) in oligonucleotide binding.

and the crosslinks constrain the aptamer's location at two points. It is clear from the model that only a small set of atoms in the molecule make direct contact with the protein target. Further, these atoms are clustered spatially along a relatively small region of the aptamer surface. These groups are probably key to the high affinity interactions of the aptamer for L-selectin protein.

Results and Discussion

Post-SELEX modification

Although the high affinity aptamers derived from SELEX experiments usually exhibit a high degree of specificity, it is possible to improve both affinity and specificity through post-SELEX modifications. Modified ligands can be prepared and screened for enhanced binding activity and/or other properties such as improved pharmacokinetics, tissue distribution, and nuclease resistance. For this discussion we note the ease with which modifications can be made in and around the binding interface, trying to identify nucleotide adducts that provide extra binding energy with the intended target protein. As an aside, we have made possible SELEX itself with modified oligonucleotides in the starting library, taking advantage of new synthetic triphosphates. A variety of chemistry is known for the modification of ribonucleosides.[29] Optimally, modified nucleotides would be made by chemistry that does not require any additional protection/deprotection steps.[30] We have worked out palladium catalyzed methods for the assembly of modified UTPs and established that a variety of substituents may be attached to the 5-position of UTP that are compatible with the enzymology of SELEX. Aromatic, alkyl, heterocycle, amino acid and other groups can be incorporated into SELEX ligands via modified UTPs.[31] These substituents can be used to fine-tune the electrostatic and hydrophobic interactions of the nucleic acids and their targets.

After SELEX ligands have been isolated, either with native or modified bases, the option exists for post-SELEX modification of bases and/or riboses. There are several examples of post-SELEX modifications that result in enhanced binding over the original sequence. The affinity of a DNA aptamer to basic fibroblast growth factor (bFGF) was improved by single-sequence site substitutions of modified thymidine bases.[32] The substitution site corresponds to a position of low sequence conservation surrounded by highly conserved residues in a tetraloop (Fig. 3a). Two substitutions were introduced, namely 5-[*N*-(aminoethyl)-3-acrylamido]deoxyuridine and 5-[*N*-(aminohexyl)-3-acrylamido]deoxyuridine, and each improved the affinity of the original aptamer, the latter by over fivefold. In another example post-SELEX modification was performed on an aptamer to vascular endothelial growth factor (VEGF) isolated from a modified oligonucleotide random library containing 2′-aminopyrimidines with the intent of increasing stability against enzymatic degradation.[33] Derived from an active parent sequence, multiple pools of 2′-*O*-methylpurine substituted aptamers were synthesized. Each pool contained only three to four sites of substitution. After affinity partitioning of each pool, sites of tolerated substitution were identified using alkaline hydrolysis and electrophoretic analysis. The combined results of these experiments led to the synthesis of a molecule in which the ribose 2′-hydroxyl was replaced with 2′-*O*-methyl in 10 out of 14 purine positions. The resulting molecule, shown in Figure 3(b), not only exhibited increased nuclease resistance but also had a 17-fold increase in binding affinity to VEGF over that of the original aptamer.

A dramatic change toward higher affinity ligands could be achieved by the introduction of either reversible or irreversible crosslinking groups. An oligonucleotide that

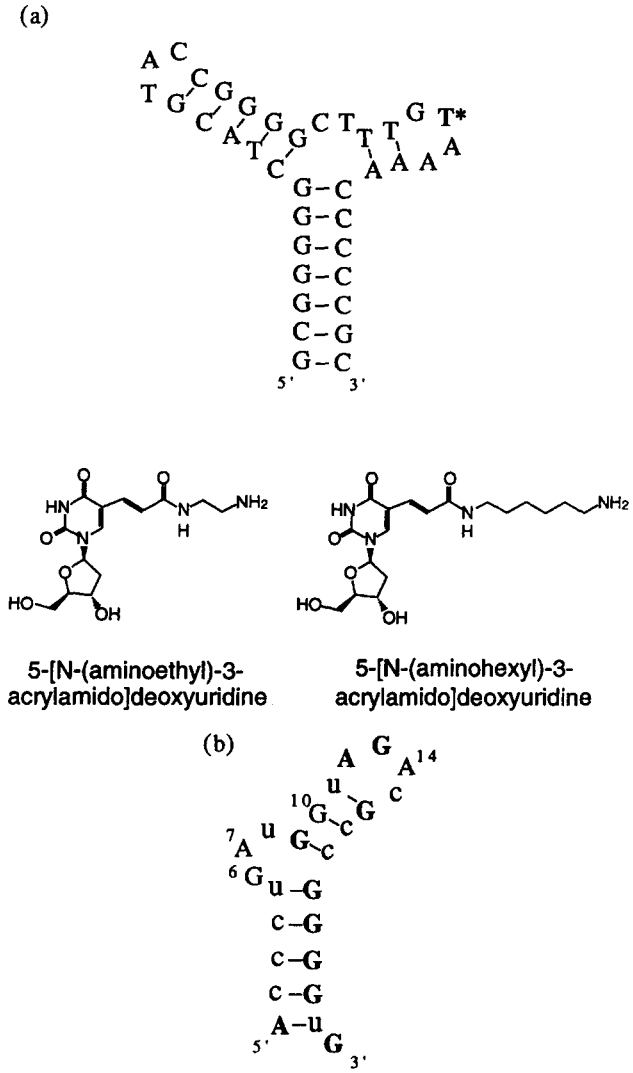

Figure 3. (a) The proposed secondary structure of the bFGF DNA aptamer used for post-SELEX modification. The two modified deoxyuridine nucleosides shown were substituted for the thymidine nucleotide denoted in bold and marked by an asterisk. (b) The proposed secondary structure of the VEGF aptamer used for post-SELEX modification. 2′-O-methyl purine nucleotides are in bold-face and the remaining ribopurines are numbered. The 2′-aminopyrimidine nucleotides are denoted in lower case.

binds tightly and specifically to a protein target is likely to have various nucleotides in proximity, when bound, to potentially reactive amino acid side-chains of the protein target. Correct substitution of an appropriately reactive nucleotide within the aptamer would allow crosslinking to the protein.[34,35] A library of oligonucleotides, each containing one or more reactive adducts could be mixed with the protein target and chemical crosslinking *demanded* for capture during partitioning. Successful crosslinking would allow the protein target and the covalently bound oligonucleotide to be partitioned by any barbaric procedure, including, for example, denaturing gel electrophoresis. Such covalent complexes could be isolated from the gel and amplified to provide the subset of oligonucleotides that crosslinked to the protein target. Exactly that protocol has been successfully developed and implemented with HIV-1 *rev* protein.[36] The resulting compounds provide high binding and clean crosslinking specificities, probably because covalent SELEX demands both binding and suitable juxtaposition of a reactive nucleotide with an appropriate amino acid.

Because SELEX is a thermodynamic process relying on competitive binding, any reversible linkage is compatible with the SELEX process as long as the kinetics of equilibrating the nonspecific crosslinks are sufficiently fast (1 h–1 day) to conduct the partitioning. For irreversible crosslinking the reaction rate must be significantly slower for the nonspecific or lower affinity binders as compared to the 'winners'. In other words, specific high affinity binding must precede the chemical reaction. In addition, irreversible crosslinking SELEX requires (as above) an amplification protocol that is compatible with the crosslinking chemistry.

Crosslinking SELEX could also be performed as a post-SELEX exercise. This would guarantee that high affinity binding would precede chemical reaction since the starting pool would have been preselected for such activity. But rather than testing single substitutions for crosslinking, a pool of molecules derived from a high-affinity sequence would be best. Indeed, for post-SELEX optimization of binding affinity with functional groups not present in the original library, a combinatorial approach would offer enormous advantages. Many more possible substitutions could be tried over many more positions. Multiple concurrent substitutions could result in activity not predicted by single events. For these reasons, a combinatorial approach to post-SELEX optimization of aptamers is being developed.

Post-SELEX combinatorial aptamer modification

Any combinatorial chemistry protocol may be generally described as three separate activities, molecular library synthesis, selection or screening, and structural deconvolution. Numerous factors must be taken into account at each step to assure convergence on the 'best' drug leads from the starting pool. In this section we will discuss combinatorial approaches for the modification of SELEX-derived aptamers. Some of the possibilities for synthetic modification, selection and structural deconvolution of modified aptamers will be described. The protocol begins with a defined sequence which is embellished with various new functionalities. The examples given only serve as a starting point for creative solutions to the post-SELEX enhancement of in vivo efficacy of aptamers.

RNA library synthesis

Clearly, there are a plethora of reactive sites on an RNA aptamer that may be elaborated by a broad array of synthetic chemistries. Modification may be performed on the heteroatoms (oxygen and nitrogen) and double bonds of the bases, hydroxyls of the ribose and the phosphodiester backbone. Our initial research has

been on synthetic methods for modifications of the pyrimidine and purine bases at specific locations and the 2′-hydroxyl of the ribose. The decision of what functionalities to include in the library can be influenced by the knowledge of what protein side-chains are in contact with the aptamer and the extent of solvent accessibility. This information can be obtained by traditional chemical footprinting and crosslinking techniques (Fig. 2c, L-selectin/aptamer model) as well as NMR spectroscopy and X-ray crystallography. Once the preferred sites of aptamer modification have been determined the question arises as to how to prepare the library.

In general, combinatorial chemistry library synthesis may be accomplished by mixed synthesis, split synthesis or parallel synthesis.[37] Both solution- and solid-phase synthesis techniques have been used. Recently, the combinatorial chemistry paradigm has shifted away from mixed synthesis of thousands of compounds to libraries containing a few hundred members because of the 'weak lead' problem. This problem results from multiple modest affinity molecules giving rise to a large positive signal, masking the activity of superior leads present at lower concentrations in other mixtures. The 'weak lead' problem is common and is a direct result of the difficulty in screening/deconvoluting mixtures. Nevertheless, for aptamers, mixed synthesis is preferred and less prone to these difficulties because of the exquisite specificity of the starting aptamer for its target and the use of stringent partitioning techniques. Here we describe a mixed synthesis solid phase approach that may be applied to aptamers to generate tens of thousands of compounds in the library.

In the library-mixed synthesis there are two options: either assemble the RNA with a mixture of modified nucleotide monomers, or treat the aptamer with various modifying reagents subsequent to oligonucleotide synthesis. As in any synthetic problem, the efficiency of the more convergent scheme is always preferred, making solid-phase synthesis of an aptamer with modified nucleotides of greater utility. In addition, chemical modification of aptamers after their synthesis can give a distorted representation of the intended library because of the relative reaction rates of the various starting reagents with the folded RNA. However, building diversity via solid-phase oligonucleotide synthesis requires stream-lined methods to prepare the modified nucleoside phosphoramidite reagents. Ideally, modification of the nucleosides would require only a single additional step in the total synthesis of the phosphoramidite. We have developed such chemistry for the modification of both pyrimidine and purine nucleosides.

Modified aptamer design

In designing the post-SELEX aptamer library it is crucial that the modifications do not disrupt the folded aptamer. For example, all aptamers have secondary structural units that require specific functional groups for hydrogen bonding. Preferred modifications do not interfere with these essential hydrogen bonding groups. For this reason we have chosen the 5-position of pyrimidines, the 8-position of purines, and the 2′-position of all nucleotides as modification sites for aptamer library synthesis. Currently, approximately 80 different modified pyrimidine or purine phosphoramidites can be made by simple techniques or are commercially available. Choosing which modifications are desirable to include in the modified aptamer library can be decided in at least two ways. The first, and least rational approach is to introduce new functional groups in a completely random way, where hydrophobic, hydrophilic and charged (both negative and positive) groups are added with the hope that some combinations will be superior to the starting aptamer. The second approach requires some basic understanding of the protein epitope where the oligonucleotide binds and the identity of the side chains on or close to the protein surface. For the most part, the decision is about what type of groups *not* to include because library size, partitioning methods and deconvolution techniques are limiting.

Aptamers that contain functional groups capable of forming specific crosslinks to protein targets are attractive. The ultimate goal is to achieve slow tunable off rates for aptamer/target dissociation. The appended groups that crosslink may also contain hydrophobic, hydrophilic or charged functionality. A plethora of modified pyrimidine and purine nucleosides with reversible crosslinking groups can be easily assembled and the rationale for their inclusion into aptamers is simple. Reversible crosslinking chemistries are easily adapted to the selection of improved aptamers via competitive equilibrium binding. However, the scope of chemistries amenable to reversible crosslinking is limited because of the reactive groups available on proteins (amino, thiol, guanadyl and hydroxyl) that will react under physiological conditions. Nevertheless, numerous options exist. Crosslinking may be accomplished by the formation of imine, acetal, ester and disulfide linkages, as well as, conjugate addition to α,β-unsaturated carbonyl linkers. Examples of 2′-deoxyuridine nucleosides that are amenable to phosphoramidite synthesis and inclusion into post-SELEX aptamer libraries are shown in Figure 4. Hydrophobic groups as small as vinyl (**1**) and as large as pyrenyl (**13** and **14**) have been successfully prepared in a single step.[38,39] Numerous crosslinking carbonyl compounds (**3**, **6**, **7**, **8**, **9**, **10**, **11** and **14**) with varying degrees of side-chain hydrophobicity have also been prepared in a single step. The reactivity of these carbonyl linkers could vary significantly depending on the aptamer–target interface.

RNA uridine nucleoside modifications are shown in Figure 5.[29] Here hydrophobic groups have been attached in the form of ketones (**17** and **18**)[30] or amides (**24** and **27**).[40] The amides add new hydrogen bonding capabilities to the aptamer. Here too, crosslinking carbonyl groups can be attached to the 5-position of

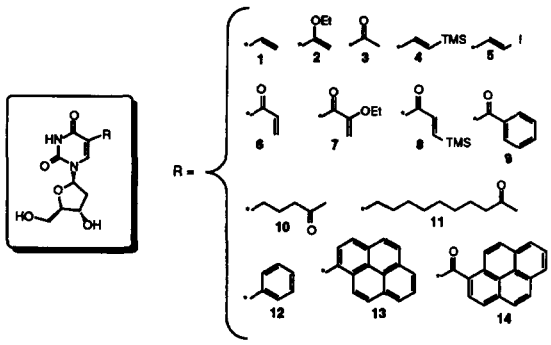

Figure 4. A sample of modified 2′-deoxyuridines amenable to post-SELEX combinatorial mixed solid-phase synthesis.

uridine (**15**, **16**, **17** and **18**). Figures 4 and 5 are only a sampling of the pyrimidine nucleosides that have been prepared and any functional group that may be appended to a DNA nucleotide can also be attached to an RNA nucleotide. It should be noted that while discussion here has been limited to 5-position modified uridines, related analogues of cytidine have been prepared so that the combinatorial modification of an aptamer can include either or both of the pyrimidine bases.

Because most SELEX-derived aptamers contain a significant fraction of purine as well as pyrimidine nucleotides, facile chemistry for the elaboration of purine nucleosides has also been developed. Some examples that reveal the breadth of the chemistry are given in Figure 6. Both DNA and RNA purine nucleosides have been modified at the 8-position with a wide array of functional groups. Simple hydrophobic substituents can be attached to the 8-position (**28**, **29** and **30**),[41] as well as varying degrees of steric bulk via an amide linkage (**31**, **33**, **34**, **37** and **38**).[42] Hydrophilic (**35**) and charged (**36** and **39**) groups may also be appended to the 8-position of purine nucleosides. Finally, even functional groups that have a known affinity for a protein target can be attached to the 8-position of purine nucleosides. Biotinylated nucleosides depicted as **40** serve as an example.

The remaining two sites available for modification of an aptamer are the 2′-position of RNA and the phosphodiester oxygens of both RNA and DNA. Here we will discuss our approach to the modification of the 2′-

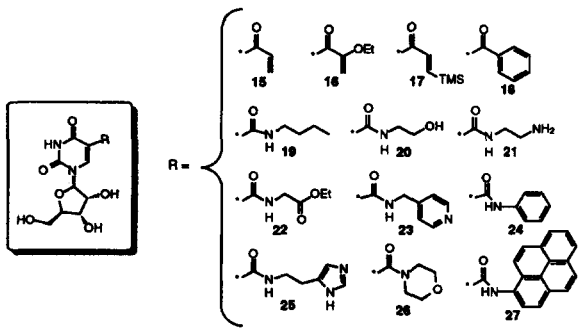

Figure 5. A sample of modified uridines amenable to post-SELEX combinatorial mixed solid-phase synthesis.

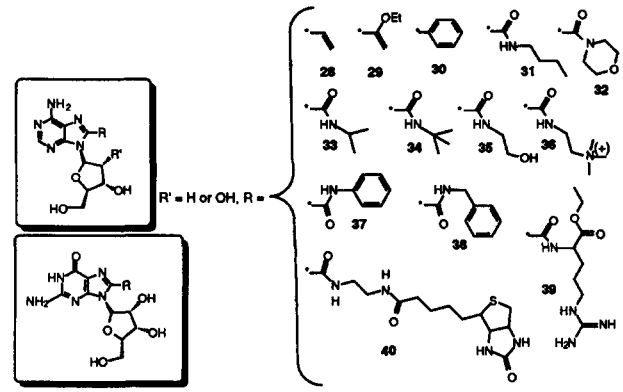

Figure 6. A sample of modified 2′-deoxyadenines, adenines and guanosines amenable to post-SELEX combinatorial mixed solid-phase synthesis.

position (Fig. 7). Streamlined stereospecific methods for the synthesis of 2′-modified pyrimidine nucleosides have been reported previously.[43–45] The clear advantage of this chemistry is that essentially any hydroxyl amine and many amines can be used to form a wide array of uridine derivatives. As with the 5-position modification discussed above, hydrophobic functional groups (**41**, **42**, **43**, **46** and **49**) with wildly different topologies can be attached to the 2′-position. Hydrophilic (**45** and **47**) and reactive crosslinking groups can also be appended. The hydrophilic fucose derivative **45** can also be viewed as a lectin-specific affinity modification, superficially similar to the biotin derivative **40**. The 2′-modification chemistry provides another attractive feature. Uridine **48** can serve as a common intermediate that may be elaborated by substitution of the imidazole with a wide variety of hydrophobic, hydrophilic, charged and potential crosslinking groups, prior to activation as the phosphoramidite reagent used in solid phase synthesis. Derivatives of **48** with varying chain lengths can also be prepared. Finally, all of the 2′-modified uridines can be readily transformed into cytidines.

The chemistry has been well developed for the facile modification of the 5 and 2′-position of pyrimidine nucleosides and 8-position of purine nucleosides. Clearly, if only 20 of the nucleosides discussed were used in a post-SELEX combinatorial aptamer synthesis

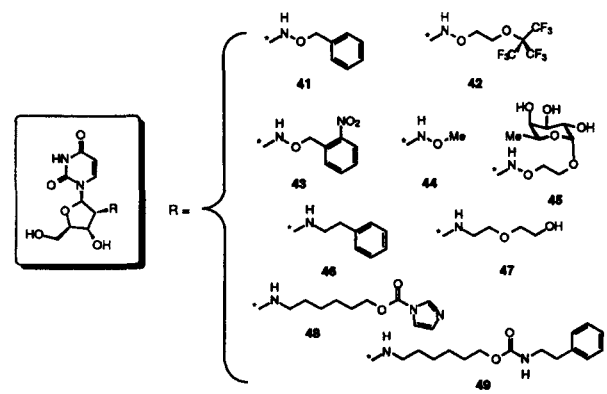

Figure 7. A sample of modified uridines amenable to post-SELEX combinatorial mixed solid-phase synthesis.

the library would be enormous. For example, consider an aptamer of only 20 nucleotides in which every base was modified. The total library would be 20^{20}, a number larger than Avogadro's number, making it impossible to get complete representation of every member of the library. To get complete coverage for all modifications of the library during the synthesis (typically 10–100 μmol) requires selection of a limited number of modifications (<10) made at each site elaborated in the sequence. Clearly, very large libraries can be prepared; these libraries require efficient and stringent partitioning techniques.

Partitioning

Partitioning, by definition, is the separation of library members superior in the selected property from the remainder of the combinatorial mixture. Partitioning can be based on a number of desired functions including enhanced in vivo efficacy, improved nuclease resistance, or altered pharmacokinetics to render the modified aptamer better suited for pharmaceutical development. Regardless of the motivation, retaining, or more often increasing, the affinity of a given aptamer for its protein target is a common objective of all post-SELEX combinatorial strategies. Equilibrium dialysis and immobilized protein methods such as affinity chromatography are both viable partitioning techniques. Where increased protein affinity is the ultimate goal equilibrium dialysis is perhaps better suited.

Equilibrium dialysis allows for binding competition between aptamers under excess oligonucleotide conditions while maintaining careful control of homogeneous protein concentration. Once partitioned, dissociation of selected aptamers from the target can be accomplished by a number of means including denaturing the protein, denaturing the oligonucleotide, reversing crosslinks or incubation with an excess of a competitive binder. Subsequently, the mixture of selected aptamers can be deconvoluted to determine the location and identity of beneficial modifications.

Deconvolution

Identification of favorable modifications (deconvolution) can be achieved by a number of means. Alkaline cleavage of hydrolytically unstable substitutions coupled with gel electrophoresis can be used to deconvolute certain libraries (see below) but is obviously limited. A similar, less restricted, approach based on nuclease stability could be envisioned. If a library was based on a modification which imparted some exonuclease stability, then advantageous changes could be identified by exonuclease incubation followed by electrophoretic analysis in a manner analogous to alkaline hydrolysis deconvolution. However, unambiguous identification could be obtained only if single modifications were employed. Multiple isotope labeling is an approach that would allow for more than one type of modified nucleotide to be introduced at a given position, but is restricted by the number of available isotopes and the enormous and tedious task of incorporating them into the combinatorial members. A much more straightforward strategy which solves the generality problem while being able to locate and identify multiple modifications relies on electrospray mass spectrometry (ESMS) for deconvolution.

Based on previous applications, ESMS could be well suited for the deconvolution of combinatorial oligonucleotide libraries. Mass spectral analysis has been demonstrated to be an effective means of analyzing small molecule combinatorial libraries. In addition, ESMS has revolutionized the sequencing of oligonucleotides[46] to such an extent that even subtle modifications of natural ribonucleotides may be identified in cellular RNAs.[47] It is therefore a logical extension that ESMS could prove to be a powerful technique for the position-specific identification of favorable aptamer modifications.

We have explored the feasibility of deconvoluting combinatorial oligonucleotide mixtures using ESMS sequencing by synthesizing a 12-mer oligonucleotide library in which three fixed positions were randomized with three thymidine analogues creating a library of 27 unique sequences. Collisional dissociation of each parent ion produced secondary spectra from which it was possible to sequence each individual oligonucleotide from the library.[48] The sequencing data obtained from ESMS can distinguish each modified base, thereby identifying each unique molecule. While this library was relatively modest in population and oligonucleotide length, it pushed the limits of current ESMS sequencing technology suggesting that deconvolution of larger libraries would be arduous. However, one can envision dissecting longer combinatorially modified aptamers into pieces of manageable sizes and sequencing them by ESMS. Cleavage positions can be generated by the judicious placement of hydrolytically labile ribose nucleotides. The following example outlines how this can be done combinatorially.

The goal of this experiment with an L-selectin aptamer was to identify those deoxyribonucleotides in the aptamer that could be substituted with a ribonucleotide while maintaining affinity comparable to that of the original L-selectin aptamer. To this end, a combinatorial oligonucleotide library was synthesized wherein each member contained approximately one ribonucleotide substitution located at any position except the 3′-terminus. Partitioning was accomplished using nitrocellulose filter binding. The selected library members were then subjected to alkaline hydrolysis, analyzed by gel electrophoresis and compared to unselected ribose-modified oligonucleotides that had been hydrolyzed in an analogous manner. The results are shown in Figure 8. The gel band pattern and the phosphorimage contour profile indicate that ribose substitution at each position in the unselected library is roughly equal. In contrast, the pattern for the partitioned oligonucleotides is noticeably different and suggests that certain ribose

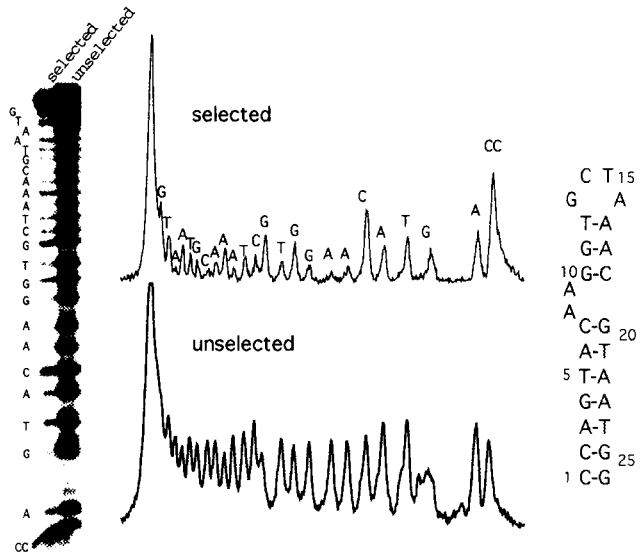

Figure 8. Gel electrophoretic analysis and phosphorimager-generated profiles of selected and unselected ribose-modified libraries. Tolerated positions are represented by darker bands in the selected lane or higher peaks in the corresponding profile. The unselected library is present as a control to demonstrate regular incorporation of ribonucleotides in the library. A proposed secondary structure of the DNA L-selectin aptamer is displayed to the right.

substitutions are detrimental to protein binding. With the information obtained, certain single ribose-modified oligonucleotides were individually synthesized and their affinity for the protein measured. All but one of the synthesized oligonucleotides representing tolerated ribose substitution displayed protein binding similar to, if not better than the original aptamer.

Experiments such as this can be used to identify cleavage sites within an aptamer library. The fragments produced on cleavage are designed to be used for ESMS deconvolution in the following manner. After partitioning, the selected library can be hydrolytically cleaved into pools of fragments of known length and parent sequence. These pools, much less complex than the original library, could be resolved by ESMS and members identified using predicted molecular weights. With such an approach, beneficial modifications within complex combinatorial aptamer libraries consisting of thousands of members could be identified via the ESMS sequencing of much smaller sub-libraries. An aptamer cleaved into M pieces, each with a complexity N, would have a library population of M^N, where N does not exceed the number of unique sequences that are resolvable using ESMS. For example, an aptamer library comprised of 390,625 unique members could be fragmented at three cleavage sites into four pools, each containing 25 distinct oligonucleotides. Each pool could be easily deconvoluted using ESMS.

In practice, partitioning can also be used to reduce the library population to any desired level. For example, competitive partitioning using a 10-fold excess of oligonucleotide to target should result in isolation of a mixture a fraction of the original library size. Such a reduction following partitioning suggests that original

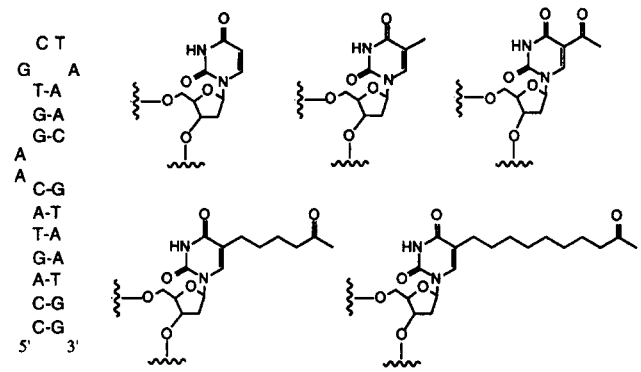

Figure 9. Secondary structure model of single-stranded DNA L-selectin aptamer and structures of deoxy pyrimidine nucleotides used to create a combinatorial library. Modified nucleosides were randomly substituted at each T position in the aptamer sequence.

library complexities considerably larger than deconvolution limits allow can be employed. Combining this reduction with that obtained through dissection allows initial library complexity to approach 10^6–10^7 unique members.

Conclusions

In closing, we describe work in progress with the L-selectin ligand illustrative of our current efforts at performing post-SELEX optimization. In particular, we are testing the feasibility of the ESMS approach for the deconvolution of post-SELEX combinatorial mixtures. A library was created by randomly substituting one of five base-modified thymidines at all T positions creating a library of 3125 members (Fig. 9). Three of the modifications presented methyl ketones to allow for schiff base formation between a given aptamer and a lysine side chain. In addition, ribo-cytidine and ribo-adenosine were incorporated at positions 7 and 17, respectively, to facilitate dissection of partitioned modified aptamers via alkaline hydrolysis. Thus, even without partitioning, the intricate library of 3125 26-mers can be disassembled into three modestly sized mixtures of five seven-mers, 25 10-mers and 25 nine-mers. Currently, ESMS deconvolution of the unselected library is being explored prior to partitioning experiments.

We have described a general strategy for post-SELEX optimization of aptamers. A strong foundation for each of three steps comprising the process has been laid. Considerable progress has been made toward the synthesis of comprehensive libraries. Partitioning is quite specific to the particular function desired for optimization, and can be developed as needed. Finally, the deconvolution step is the primary focus of ongoing research. Nevertheless, we are at a sufficiently advanced stage where all three steps are being combined to pursue challenging combinatorial post-SELEX optimization experiments.

Experimental

Ribonucleotide-substituted L-selectin aptamer library

The oligonucleotide library was synthesized on an ABI 392 DNA synthesizer using standard DNA cycles and conditions except as indicated below. Reagents, including phosphoramidites, were obtained from Glen Research. All syntheses were performed at the 1 μmol scale using 500 Å controlled pore glass (CPG) support from Millipore with the final dimethoxytrityl (DMT) protecting group removed prior to oligonucleotide cleavage from the resin. The ribonucleoside-doped library was prepared using appropriate phosphoramidite reagents consisting of the deoxyribonucleoside (50 mM) and the corresponding ribonucleoside (50 mM). U rather than riboT was used in all dT positions. Ribonucleotides were substituted at each position except for the 3′-terminal residue. The oligonucleotide library was cleaved from the CPG support and deprotected by incubation at 55 °C for 14 h in 1 mL of 6 M NH_3/ethanol. Following filtration and concentration to dryness, the library was further deprotected by treating with 2 mL of 1 M tetratbutyl ammonium fluoride for 48 h at ambient temperature in the dark. The oligonucleotide mixture was then desalted by gel filtration through two Nap columns (Pharmacia) using 100 mM triethylammonium acetate (TEAA) pH 7 as the eluent. After concentration to dryness, the sample was used directly in the binding experiments described below.

Ribonucleotide-library binding to L-selectin and identification of tolerated sites of modification

The library was gel-purified and 5′-end labeled by standard methods. The library's bulk K_d for L-selectin was determined by nitrocellulose filter binding to be 60 nM. In a 50 μL volume, 1×10^6 cpm of the library, at 10 nM, was incubated with 20 nM L-selectin in binding buffer (20 mM HEPES pH 7.5, 125 mM NaCl, 1 mM $MgCl_2$, 1 mM $CaCl_2$, 5 mM KCl, 0.01% (w/v) human serum albumin) at 37 °C for 15 min. Bound nucleic acid was recovered by filtration across a nitrocellulose filter (HAWP 013, Millipore). Nucleic acid was extracted from the filter by the addition of 300 μL, 5 M urea, and 300 μL phenol. After 1 h at 22 °C, 200 μL ddH_2O and 200 μL chloroform were added, the tube was vortexed and centrifuged, and the aqueous phase removed. Five μg tRNA was added as a carrier and the nucleic acid was precipitated with ethanol. The unbound library sample was also subjected to the phenol/urea extraction and precipitation procedure. For alkaline hydrolysis at ribonucleotide positions, dried precipitate was resuspended in ddH_2O, brought to 50 mM Na_2CO_3 pH 9.0/1 mM EDTA and incubated at 95 °C for 2 h. Formamide was added to 50% and the samples separated on a 20% polyacrylamide/7Murea/1 × TBE gel, which was visualized using a Fujix BAS 1000 phosphorimager. Size markers were created by alkaline hydrolysis of the identical DNA sequence with five defined positions of ribonucleotide substitution.

Modified deoxynucleotide-substituted L-selectin aptamer library

The oligonucleotide library was synthesized on an ABI 392 DNA synthesizer using phenoxyacetyl (Pac) protected phosphoramidites and standard DNA cycles and conditions except as indicated below. Reagents, including unmodified phosphoramidites, were obtained from Glen Research. All syntheses were performed at the 1 μmol scale using 500 Å CPG support from Millipore with the final DMT protecting group retained prior to oligonucleotide cleavage from the resin. The oligonucleotide library was synthesized by using a modified nucleoside phosphoramidite mixture (0.20 mM solution of each of the pyrimidine bases shown in Fig. 9) at each dT position. In addition, ribonucleoside phosphoramidites were substituted at positions seven and 17. Standard RNA synthetic cycles were used at all dT and ribonucleoside positions. The library was cleaved from the resin and deprotected by incubation in 1.2 mL of 75% concentrated NH_4OH/ethanol for 48 h at room temperature. Removal of the silyl protecting groups and desalting were carried out as described above for the ribonucleotide modified library. The DMT-protected library was purified using a Rainin Dynamax HPLC equipped with a Rainin UV D-II dual wavelength detector. A Vydac C4 reverse phase preparatory HPLC column was used at a flow rate of 9 mL/min with 100 mM TEAA pH 7.0 and acetonitrile (MeCN) as eluents. A 15–40% acetonitrile gradient over 20 min was employed. Fractions eluting between 10.5 and 16.5 min were collected and concentrated to dryness. The 5′-DMT protecting group was removed by treating the oligos with 80% acetic acid at ambient temperature for 1 h, after which time the sample was concentrated to dryness under reduced pressure. The fully deprotected oligos were HPLC purified using a 5–30% CH_3CN gradient over 20 min. Fractions eluting between 8 and 16 min were collected and concentrated to dryness.

Acknowledgements

We were fortunate to be able to talk to a large group of scientists at NeXstar Pharmaceuticals, Inc. and the Boulder campus of The University of Colorado. In particular we would like to thank David Parma and Chi Tu.

References

1. Tuerk, C.; Gold, L. *Science* **1990**, *249*, 505.
2. Houghton, R. A.; Pinilla, C.; Blondelle, S. E.; Appel, J. R.; Dooley, C. T.; Cuervo, J. H. *Nature (London)* **1991**, *354*, 84.
3. Martin, E. J.; Blaney, J. M.; Siani, M. A.; Spellmeyer, D. C.; Wong, A. K.; Moos, W. H. *J. Med. Chem.* **1995**, *38*, 1431.
4. Smith, G. P.; Scott, J. K. *Methods Enzymol.* **1993**, *217*, 228.

5. Irvine, D.; Tuerk, C.; Gold, L. *J. Mol. Biol.* **1991**, *222,* 739.

6. Lorsch, J. R.; Szostak, J. W. *Biochemistry* **1994**, *33,* 973.

7. Gold, L.; Polisky, B.; Uhlenbeck, O.; Yarus, M. *Annu. Rev. Biochem.* **1995**, *64,* 763.

8. Shoichet, B. K.; Baase, W. A.; Kuroki, R.; Mathews, B. W. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92,* 452.

9. Bartel, D. P.; Szostak, J. W. *Science* **1993**, *261,* 1411.

10. Seed, B. *Chem. Biol.* **1994**, *1,* 125.

11. Austin, D. J.; Crabtree, G. R.; Schreiber, S. L. *Chem. Biol.* **1994**, *1,* 131.

12. Gold, L. *J. Biol. Chem.* **1995**, *270,* 13581.

13. Hentze, M. W. *Trends Biochem. Sci.* **1994**, *19,* 101.

14. Gold, L.; Brown, D.; He, Y.; Shtatland, T.; Singer, B. S.; Wu, Y. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 59.

15. Conrad, R. C.; Baskerville, S.; Ellington, A. D. *Molecular Diversity* **1995**, *1,* 69.

16. Fox, G. E.; Woese, C. R. *Nature* **1975**, *256,* 505.

17. Woese, C. R.; Pace, N. R. In *Probing RNA Structure, Function, and History by Comparative Analysis*; Gespeland, A.; Atkins, J., Eds.; Cold Spring Harbor Laboratory: Plainview, 1993; p 91.

18. Mutagenesis can be added to the protocol during amplification.

19. Davis, J.; Janjić, N.; Pribnow, D.; Zichi, D. A. *Nucleic Acids Res.* **1995**, 4471.

20. Davis, J. P.; Janjić, N.; Javornik, B. E.; Zichi, D. A. *Methods Enzymol.* **1996**, *267,* 302.

21. Fan, P.; Suri, A. K.; Fiala, R.; Live, D.; Patel, D. J. *J. Mol. Biol.* **1996**, *258,* 480.

22. Jiang, F.; Kumar, R. A.; Jones, R. A.; Patel, D. J. *Nature (London)* **1996**, *382,* 183.

23. Dieckmann, T.; Suzuki, E.; Nakamura, G. K.; Feigon, J. *RNA* **1996**, *2,* 628.

24. Yang, Y.; Kochoyan, M.; Burgstaller, P.; Westhof, E.; Famulok, M. *Science* **1996**, *272,* 1343.

25. We often focus on randomized oligonucleotide lengths that will provide a ligand short enough to synthesize *chemically* for future drug development; for gene therapy applications such restraint is not necessary.

26. Ehresmann, C.; Baudin, F.; Mougel, M.; Romby, P.; Ebel, J.-P.; Ehresmann, B. *Nucleic Acids Res.* **1987**, *15,* 9109.

27. Breiner, K. M.; Daugherty, M. A.; Oas, T. G.; Thorp, H. H. *J. Am. Chem. Soc.* **1995**, *117,* 11673.

28. Hicke, B. J., unpublished results.

29. Eaton, B. E.; Pieken, W. A. *Annu. Rev. Biochem.* **1995**, *64,* 837.

30. Dewey, T. M.; Mundt, A. A.; Crouch, G. J.; Zyzniewski, M. C.; Eaton, B. E. *J. Am. Chem. Soc.* **1995**, *117,* 8474.

31. Dewey, T. M.; Creighton, S.; Wayland, S. R.; Zyzniewski, M. C.; Eaton, B. E., unpublished results.

32. Willis, M. C., unpublished results.

33. Green, L. S.; Jellinek, D.; Bell, C.; Beebe, L. A.; Feistner, B. D.; Gill, S. C.; Jucker, F. M.; Janjić, N. *Chem. Biol.* **1995**, *2,* 683.

34. Willis, M. C.; Hicke, B. J.; Uhlenbeck, O. C.; Cech, T. R.; Koch, T. H. *Science* **1993**, *262,* 1255.

35. Hicke, B. J.; Willis, M. C.; Koch, T. H.; Cech, T. R. *Biochemistry* **1994**, *33,* 3364.

36. Jensen, K. B.; Atkinson, B. L.; Willis, M. C.; Koch, T. H.; Gold, L. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 12220.

37. Gordon, E. M.; Barrett, R. W.; Dower, W. J.; Fodor, S. P. A.; Gallop, M. A. *J. Med. Chem.* **1994**, *37,* 1385.

38. Crouch, G. J.; Eaton, B. E. *Nucleosides Nucleotides* **1994**, *13,* 939.

39. Netzel, T. L.; Zhao, M.; Kambiz Nafisi, K.; Headrick, J.; Sigman, M. S.; Eaton, B. E. *J. Am. Chem. Soc.* **1995**, *117,* 9119.

40. Dewey, T. M.; Zyzniewski, C.; Eaton, B. E. *Nucleosides Nucleotides* **1996**, *15,* 1611.

41. Tu, C.; Keane, C.; Eaton, B. E. *Nucleosides Nucleotides* **1995**, *14,* 1631.

42. Tu, C.; Keane, C.; Eaton, B. E., unpublished results.

43. McGee, D. P. C.; Sebesta, D. P.; O'Rourke, S. S.; Martinez, R. L.; Jung, M. E.; Pieken, W. A. *Tetrahedron Lett.* **1996**, *37,* 1995.

44. Sebesta, D. P.; O'Rourke, S. S.; Martinez, R. L.; Pieken, W. A.; McGee, D. P. C. *Tetrahedron* **1996**, *52*, 14385.

45. McGee, D. P. C.; Vaughn-Settle, A.; Vargeese, C.; Zhai, Y. *J. Org. Chem.* **1996**, *61,* 781.

46. Ni, J.; Pomerantz, S. C.; Rozenski, J.; Zhang, Y.; McCloskey, J. A. *Anal. Chem.* **1996**, *68,* 1989.

47. Limbach, P. A.; Crain, P. F.; McCloskey, J. A. *Nucleic Acids Res.* **1994**, *22,* 2183.

48. Pomerantz, S. C.; McCloskey, J. A.; Tarasow, T. M.; Eaton, B. E. *J. Am. Chem. Soc.*, **1997**, *119*, 3861.

(*Received in U.S.A.* 31 *October* 1996; *accepted* 18 *February* 1997)